**Pokok Bahasan V**

**K-Nearest Neighbor (KNN).**

**Kode Pokok Bahasan**: TIK.RPL03.004.00.01

**Deskripsi Pokok Bahasan**:

Membahas bagaimana penerapan K-Nearest Neighbor untuk melakukan Klasifikasi.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Elemen Kompetensi | Indikator Kinerja | Jml Jam | Hal |
| 1. | Menampilkan hasil Klasifikasi dari kasus yang diberikan. | Mampu melakukan analisis terhadap klasifikasi data dari diagram yang muncul. | 1 |  |
| 2. | Melakukan visualisasi hasil KNN menggunakan library ggplot. | Mampu melakukan visualisasi hasil KNN menggunakan library ggplot pada dataset yang ditentukan | 1 |  |

**TUGAS PENDAHULUAN**

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum :

1. Menginstal R pada PC masing-masing praktikan.

2. Menginstal R Studio pada PC masing-masing praktikan.

3. Menginstal aplikasi pengolah data (Excel).

**DAFTAR PERTANYAAN**

1. Apa itu algoritma K-Nearest Neighbor?

2. Apa kegunaan K-Nearest Neighbor?

3. Sebutkan tahapan dari proses algoritma K-Nearest Neighbor!

1. sebuah algoritma untuk melakukan klasifikasi terhadap objek dengan data pembe- lajaran yang jaraknya paling dekat dengan objek tersebut.

2. digunakan untuk proses yang disebut missing data imputation yang memperkirakan nilai-nilai yang hilang.

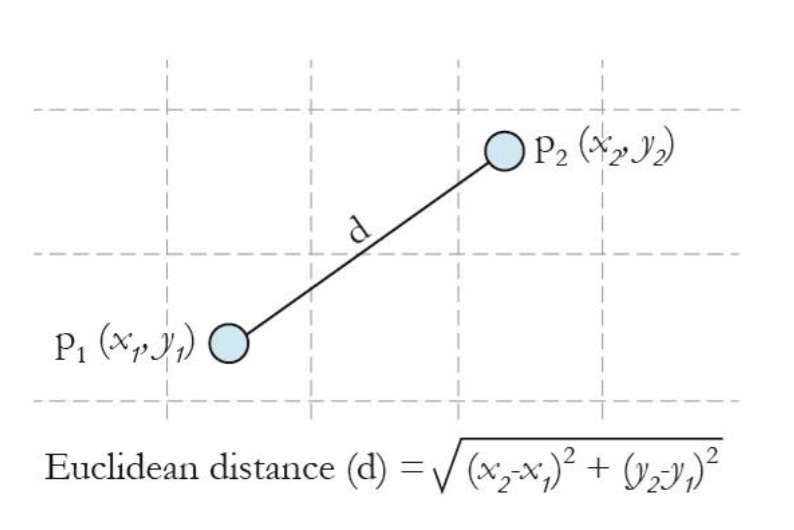
3. Tentukan jumlah tetangga (K) yang akan digunakan untuk pertimbangan penentuan kelas. Hitung jarak dari data baru ke masing-masing data point di dataset. Ambil sejumlah K data dengan jarak terdekat, kemudian tentukan kelas dari data baru tersebut.

**TEORI SINGKAT**

Algoritma Nearest Neighbor Retrieval (K-Nearest Neighbor atau K-NN) adalah sebuah algoritma untuk melakukan klasifikasi terhadap objek dengan data pembe- lajaran yang jaraknya paling dekat dengan objek tersebut.

Alasan pemilihan Algoritma KNN adalah berdasarkan data yang digunakan yaitu menggunakan data sekunder dan tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training sample, tidak memerlukan model algoritma seperti yang dihasilkan oleh algoritma lain.

KNN, Dapat digunakan untuk tujuan klasifikasi, Tidak menyusun model atau mengekstrak aturan logika tertentu sebagai hasil dari analisis, Identikasi k buah individu tetangga terdekat dilakukan dengan terlebih dahulu menghitung jarak dari individu yang akan diduga dengan setiap individu yang ada pada gugus data training. Jika ini sudah dilakukan maka tinggal mencari k buah amatan yang jaraknya paling kecil. Penghitungan jarak dari dua amatan A dan B dapat menggunakan formula Euclid distance.



**LAB SETUP**

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

1. Menginstall library yang dibutuhkan untuk mengerjakan modul.

2. Menjalankan R Studio.

3. Aplkasi pengolah data (Excel)

**ELEMEN KOMPETENSI I**

**Deskripsi:**

Menampilkan hasil klasifikasi dari kasus yang diberikan.

**Kompetensi Dasar**:

Menampilkan hasil Klasifikasi dari kasus yang diberikan.

**Latihan 1.1.1**

**Penjelasan Singkat :**

Pada latihan ini anda akan diminta melakukan analisis terhadap klasifikasi data dari diagram yang muncul.

**Langkah-Langkah Praktikum:**

1. Siapkan data : glass.csv (bisa diunduh pada link : <https://www.kaggle.com/datasets/uciml/glass> )
2. Buka Rstudio
3. Melakukan data preparation

> dataKNNGading = read.csv(“C:/Users/Sectio/Downloads/glass.csv”)

> data.training <- dataKNNGading[,1:9]

> kelas <- as.factor(dataKNNGading[,10])

> View(dataKNNGading)

> str(dataKNNGading)

Output:

|  |
| --- |
|  |

1. Membuat data amatan dengan knn

> kelas <- as.factor(dataKNNGading[,10])

> amatan.baru <- c(1.51, 14.12, 1.78, 1.79, 73.1, 0.0, 8.70, 0.76, 0.0)

> library(class)

> knn(data.training, amatan.baru, kelas, k=1)

Output:

|  |
| --- |
|  |

1. Membuat data amatan baru

> amatan.baru2 <- c(1.51, 13.12, 1.78, 1.0, 73.1, 0.2, 7.70, 0.76, 0.0)

> knn(data.training, amatan.baru2, kelas, k=5)

Output:

|  |
| --- |
|  |

**ELEMEN KOMPETENSI II**

**Deskripsi:**

Menampilkan hasil klasifikasi dari kasus yang diberikan berdasarkan dataset iris.

**Kompetensi Dasar**:

Menampilkan hasil Klasifikasi dari kasus yang diberikan.

**Latihan 1.2.1**

**Penjelasan Singkat :**

Pada latihan ini anda akan diminta melakukan analisis terhadap klasifikasi data dari diagram yang muncul.

1. Membuat plot sebelum melakukan Proses KNN

> dataku=iris

> library(ggplot2)

> ggplot(dataku, aes(x=Petal.Length, y=Petal.Width)) +

geom\_point()

Output:

|  |
| --- |
|  |

1. Melakukan pengamanatan pada data

> kelas <- as.factor(dataku[,4])

> pengamatan.baru <- c(3.8, 1.5)

> library(class)

> data.training <- dataku[,3:4]

> kelas <- as.factor(dataku[,5])

> knn(data.training, pengamatan.baru, kelas, k=5)

Output:

|  |
| --- |
|  |

1. Buat fungsi Rscript datagambar

*x <- seq(1.00, 6.00, by=0.2)*

*y <- seq(0.00, 2.50, by=0.05)*

*grid <- NULL*

*for (i in x) {*

*for (j in y) {*

*grid <- rbind(grid, c(i, j))}}*

*datagambar <- data.frame(grid)*

Output:

|  |
| --- |
|  |

1. Membuat plot dari hasil knn

> qplot(datagambar$X1, datagambar$X2)

> prediksi <- knn(data.training, datagambar, kelas, k = 5)

> plot (datagambar[,1], datagambar[,2],col=ifelse(prediksi == "versicolor", "cyan", "orange"), pch=ifelse(prediksi == "versicolor", 6, 5))

Output:

|  |
| --- |
|  |

**Tugas**

**Gunakan algorima k-NN pada R untuk kasus berikut ini dengan k=3 :**

|  |  |  |
| --- | --- | --- |
| sepal.length | sepal.width | species |
| 5.4 | 3.7 | setosa |
| 7.2 | 3.2 | virginica |
| 5.4 | 3.7 | setosa |
| 5.5 | 3.3 | setosa |
| 5.4 | 3.9 | setosa |
| 7.1 | 2.8 | virginica |
| 6.1 | 2.8 | versicolor |
| 7.2 | 2.9 | virginica |
| 6.1 | 2.7 | versicolor |
| 5.8 | 2.8 | virginica |
| 6.9 | 2.3 | versicolor |
| 5.1 | 2.7 | versicolor |
| 6.3 | 2.4 | versicolor |
| 5.4 | 2.3 | versicolor |

Data baru:

|  |  |  |
| --- | --- | --- |
| sepal.length | sepal.width | species |
| 4.2 | 3.7 | ? |

Script R :

|  |
| --- |
| >dataKNNibnu = read.csv("D:/File Kuliah Semester 5/Penambangan Data/Prak-5/tugasprak5.csv", sep=";")  >dataKNNibnu  >View(dataKNNibnu)  >str(dataKNNibnu)  >library(class)  >kelas <- as.factor(dataKNNibnu[,3])  >data.training <- dataKNNibnu[,1:2]  >amatan.baru <- c(4.2, 3.7)  >knn(data.training, amatan.baru, kelas, k=3) |

Output :

|  |
| --- |
|  |

**CEK LIST**

|  |  |  |  |
| --- | --- | --- | --- |
| Elemen Kompetensi | No Latihan | Penyelesaian | |
| Selesai | Tidak selesai |
| 1 | 1.1.1 |  |  |
| 2 | 1.2.1 |  |  |

**FORM UMPAN BALIK**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Elemen Kompetensi** | **Tingkat Kesulitan** | | | **Tingkat Ketertarikan** | | | **Waktu Penyelesaian dalam menit** |
| Menampilkan hasil Klasifikasi dari kasus yang diberikan.  yang diberikan. |  |  |  |  |  |  |  |
|  |  | Sangat Mudah |  |  | Tidak Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Mudah |  |  | Cukup Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | * Biasa |  |  | Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sulit |  |  | * Sangat Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sangat Sulit |  |  |  |  |
|  |  |  |  |  |  |  |
| Melakukan perhitungan manual menggunakan excel. |  |  |  |  |  |  |  |
|  |  | Sangat Mudah |  |  | Tidak Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Mudah |  |  | Cukup Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | * Biasa |  |  | Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sulit |  |  | * Sangat Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sangat Sulit |  |  |  |  |
|  |  |  |  |  |  |  |